

Consideraciones para construir un tesoro sobre leyes político-electorales usando tecnología digital

Catalina Naumis Peña
Universidad Nacional Autónoma de México
naumis@unam.mx

Ariel Antonio Morán Reyes
Universidad Nacional Autónoma de México
a.moran@comunidad.unam.mx

Rosa Atzimba Morales Monroy
Universidad Nacional Autónoma de México
rosa.morales@ife.org.mx

Introducción

El tesoro ha ido evolucionando como una herramienta de recuperación de información inestimable en los procesos de organización del conocimiento. Las designaciones terminológicas usadas para definir esa evolución no tienen un significado clarificador. En este caso la informática como disciplina que ha otorgado una funcionalidad diferente al tesoro ha impuesto su terminología, creando bastante confusión. En general, la difusión mayor para distinguir las dos grandes etapas han sido para los siguientes términos: tesauros terminológicos y tesauros conceptuales. Planteado así parecería que con la primera designación los tesauros no trabajarían con conceptos. Nada más alejado de la realidad porque en la organización del conocimiento se ha buscado desde un principio, la representación del contenido documental mediante conceptos y no los términos presentes en los documentos. A pesar de la opinión de muchos de los autores citados a lo largo de este trabajo se utilizará el término tesauros documentales, en lugar de tesauros terminológicos. Cesteurs, Smith y Goldberg critican la denominación de “entidades conceptuales” en el uso de los tesauros y, aún más, definirlos como “entidades abstractas que representan conceptos” (2005: 503).

No se deja de reconocer el beneficio que ha otorgado a los sistemas de información el manejo tecnológico del tesoro documental. Su desarrollo se ha ido adecuando paulatinamente a los cambios que las tecnologías de la información suponen, aunque su aplicación en la Web ha resultado hasta ahora poco fructífera. Muchos

trabajos están estudiando la relación de la semántica y las tecnologías, principalmente las representaciones del espacio semántico.

El espacio semántico es el medio en el que se involucran mínimo dos personas, las cuales pueden tener diferentes percepciones ya sea de lengua, cultura, posición política, relación familiar, nivel educativo, entre otros, por lo que es un espacio de intercambio de valores y significados, que inciden en las conductas informacionales. Por lo tanto es muy difícil transmitir el mensaje con claridad y precisión debido a que en el proceso del emisor al receptor pueden existir muchas interferencias que hagan más complicada y un poco distorsionada la comunicación. Desde una perspectiva pragmática, en la actualidad encontramos múltiples modelos y propuestas en que se aplican estos principios. Como muestra está la recuperación de información en Internet a partir de motores de búsqueda. En este caso, la inferencia en un sistema y un espacio semántico consistirían en vincular la información que se encuentra latente en la red y en las bases de datos con las necesidades de información del usuario de manera que se logre establecer la mayor precisión y relevancia en la búsqueda realizada. A esto se le denomina indización semántica latente (Erk, 2009: 106).

Los tesauros se pueden utilizar en diferentes situaciones o escenarios de aplicación, aunque en general cumplen la función de homogeneizar terminologías, controlar vocabularios o auxiliar al usuario para navegar por un espacio de conceptos. A pesar de la larga trayectoria de investigación acerca del mantenimiento de la calidad de los tesauros, hasta este momento se ha prestado poca atención a la interacción entre las especificidades estructurales de un tesoro y la calidad de los resultados en relación con los diferentes escenarios de aplicación de un tesoro. Si bien hay varias iniciativas que se centran en la calidad de los tesauros, de los metadatos (en términos de expresividad) y en la solidez estructural.

La Web semántica supone una oportunidad para utilizar los tesauros en los procesos de descripción y recuperación de información. Para muestra de ello existen múltiples iniciativas de representación de tesauros mediante RDF (Resource Description Framework)/XML (Extensible Markup Language), que han dado lugar al desarrollo de SKOS (Simple Knowledge Information System), un modelo para esquemas conceptuales aplicables en la Web semántica y que actualmente se encuentra en desarrollo por el World Wide Web Consortium (W3C). La organización del conocimiento más novedosa tiene por objeto combinar diferentes sistemas de organización del conocimiento en Internet mediante metadatos compartidos y

ontologías formales que funcionan a partir de los principios de la lógica deductiva. Inicialmente, se diseñaron modelos que seguían la lógica conceptual clásica que seguía los planteamientos neopositivistas, por lo que buscó utilizar un metalenguaje, o sea, un conjunto de instrumentos simbólicos y conceptuales independientes de las lenguas naturales para conseguir una comunicación sin ambigüedad en temas especializados.

Es necesario puntualizar que el desarrollo de un vocabulario RDF/XML para representar tesauros supone explotar de manera innovadora el modelo canónico que nos proponen las normas sobre construcción y mantenimiento de tesauros: internacionales (ISO 2788 y 5964), norteamericana (ANSI/NISO Z39.19), británica (BS 8723) y la institucional de la IFLA, vertida en la Guidelines for Multilingual Thesauri. Cabe mencionar que la norteamericana sólo contempla vocabularios controlados monolingües y que la británica incluye estándares sobre control de autoridades y una norma sobre mapas temáticos. Estas normas describen un tesoro como un conjunto de términos de diverso tipo entre los que se establecen relaciones semánticas, en un contexto que le otorga sentido. No obstante, la Web precisa de una visión en la que se comparta información, como por ejemplo que las jerarquías incluyan instancias, atributos y meronimias distinguidos entre sí, otro de los servicios a destacar es una mayor riqueza en las asociaciones asociativas. Existen, por lo tanto, múltiples iniciativas dirigidas a la definición de vocabularios bajo esquemas de RDF para sustentar tesauros.

112

Pese a que algunos autores hablan de los tesauros basados en la gestión de datos enlazados, abiertos y cerrados, lo anterior se traduce, como ya se explicó en el manejo de dos modelos básicos para la estructuración de tesauros: uno documental de relaciones en un espacio semántico (el clásico) y otros denominados tesauros conceptuales o informatizados porque existe:

[...] una navegación directamente en el documento, en una red semántica conceptual por la que se ingresa de términos más genéricos de una faceta hacia los más específicos, e inversamente (navegación vertical), que a su vez permite la transición de una clase hacia otra y de un campo de la ciencia hacia otro mediante las relaciones asociativas (navegación horizontal) por nudos polijerárquicos basados en conceptos (Deweze, 1998. p. 48).

En este segundo tipo de tesauros que son de carácter automatizado en la construcción y en la recuperación de información se aprecia un incremento pragmático de las relaciones en especial las de asociación, y se reflejan también ideas combinadas que se vinculan entre sí con una red de relaciones semánticas de distinto tipo, principalmente jerárquicas (término), asociativas (termino relacionado) y de equivalencia (término equivalente, término preferente) (Tudhope, Harith & Jones,

2001). En esta segunda alternativa se concibe al tesauro como un conjunto de conceptos entre los que se establecen relaciones semánticas similares a las que se establecen en los tesauros documentales. Los conceptos se vinculan con términos, a modo de etiquetas léxicas, entre los que pueden o no establecer relaciones léxicas que es la propuesta del modelo SKOS.

La red semántica establecida por los tesauros conceptuales no se agota en la explicitación de las relaciones conceptuales anteriormente mencionadas, sino que incluye las denominadas notas de alcance donde se acota el significado o alcance conceptual que es representado por un término en particular. Sacco y Bothorel arguyen al respecto que:

El objetivo usual de un tesauro es guiar al indizador y al especialista en recuperación de información para escoger un mismo término para un mismo concepto. Los conceptos son representados por términos, y para cada concepto es seleccionada una de las representaciones posibles como el término preferido (2010: 8).

Los tesauros documentales posibilitan la representación de esquemas más compactos y cercanos al modelo tradicional, pero los tesauros basados en conceptos relacionados automáticamente, que podríamos denominar, como se propuso anteriormente, tesauros informatizados permiten navegar en las estructuras de información, aunque en un principio su comprensión no resulte tan intuitiva como los primeros. (Cfr. Jones, 1993). La estructuración de los tesauros documentales se limita a los tipos de relaciones que se encuentran en los mismos términos. Además, en los tesauros informatizados está presente la elección de los descriptores (o sea, los términos que representan a los conceptos), entendiendo que es una elección de forma, es decir, que los descriptores sean compuestos (formados por varios conceptos) o simples (formados por un único concepto). No obstante, un primer problema que se presenta con los tesauros informatizados es la necesidad de dotar a los usuarios de estructuras de conocimiento que le permitan un acceso rápido y amigable a la información. Los usuarios requieren de herramientas conceptuales y semánticas destinadas a una efectiva organización de la información digital, y los tesauros son una de las estructuras que pueden tener una participación activa en este sentido.

El cambio de paradigma en la modelización de tesauros supone un gran esfuerzo aunque dicho cambio se ve compensado con creces al optimizarse los procesos de elaboración, mantenimiento y productos derivados de la aplicación del tesauro en la indización de documentos (Pastor Sánchez, Martínez Méndez & Rodríguez Muñoz,

2009). La principal razón para este cambio se debe a que los tesauros documentales conllevan problemas de aplicación en procesos de indización cuando se realizan operaciones de cambio de términos preferentes y términos no preferentes. De esta manera, cuando un descriptor con el que se ha indizado un documento pasa a ser un no descriptor hay que realizar una reorganización de los índices que asocian dicho descriptor con los documentos. Hoy día es más que necesario contar con una visión amplia en la elaboración de herramientas conceptuales, sin circunscribirnos al modelo clásico de construcción de tesauros, sino considerando otros tipos de estructuraciones conceptuales como las taxonomías o las ontologías. Una ontología como recurso lingüístico tiene claros elementos de contacto con la construcción de herramientas documentales de indización y recuperación de información, ya que ofrece una estructuración formal y más eficaz de la información. Pastor Sánchez define el tesoro actual como una especie del género ontologías.

1. Los Sistemas de Información Jurídica

La información jurídica por su naturaleza posee peculiaridades determinantes para influir en el planeamiento de su organización. El hecho de que la información mencionada tenga un papel relevante radica en que el mensaje legal está integrado de un conjunto de expresiones normativas, ordenadas para indicar acciones humanas en un contexto social (global, general e inmediato) y formuladas por medio de signos que, generalmente, no van de forma definida del remitente (que son los legisladores) al receptor (que es la mayoría de las veces un individuo desconocido específicamente). La naturaleza de la documentación jurídica no hace fácil la tarea de elaborar un tesoro o cualquier tarea de control de vocabulario, ya que los descriptores evocan ideas categóricas, bajo las cuales se repiten las mismas clases en diferentes categorías y no un contexto lingüístico determinado, es decir su naturaleza está enmarcada en un ámbito socio-cultural-político con el lenguaje jurídico.

A diferencia de la mayoría de las otras disciplinas, el derecho no utiliza un lenguaje puramente descriptivo o preciso. El lenguaje jurídico no describe al derecho como objeto, sino en gran medida al derecho mismo; por otra parte, cada uno de los términos empleados no corresponde biunívocamente a una realidad o a un objeto: la ambivalencia estriba en la naturaleza profunda del derecho, a lo cual difícilmente se presenta objeción. Las palabras que integran los documentos jurídicos no pueden ser

consideradas unidades fundamentales con un sentido cada una. Así, cada documento se caracteriza de manera única por las palabras utilizadas (a causa del estilo, de la sintaxis y del vocabulario jurídico) y se emplean esos términos en su sentido más general. El estilo jurídico recurre con frecuencia a comparaciones o efectos que pueden hacer intervenir una noción análoga en el curso de un texto; pero también puede sugerir una noción o situación sin que el término que señala más habitualmente aparezca en realidad.

Ante el proyecto de desarrollo de un tesoro político-electoral, es necesario precisar que la salvedad descrita anteriormente vuelve sumamente compleja la creación de un sistema jurídico documental, y mucho más para un lenguaje controlado especializado en política electoral del sistema jurídico, ya que estaría compuesto de mensajes normativos vigentes y en constante cambio, que tienen que ser recuperados cuando sea pertinente a la necesidad del usuario, cuya situación se encuentra controlada por la norma (Rosales Salinas, 2012: 113).

Otra peculiaridad de la información jurídica radica en el hecho de que un concepto legal posee un contenido prescriptivo, de tal suerte que la función social de una ley o decreto emanado de la actividad parlamentaria requiere que tal mensaje sea reconocido por el receptor para que la pueda aplicar. Es decir, uno de los mayores problemas de los lenguajes jurídicos no es de carácter técnico sino de índole social (Perales Ojeda, 1981: 37). En toda búsqueda documental, incluso la más elemental, encontramos documentos que no son relevantes para el tema y, por el otro lado, documentos que siendo de una gran significación no son descubiertos, para lo cual se necesita un esquema de representación adecuado que combata “accidentes lingüísticos” como la polisemia (Matute Cifuentes, 1989: 111). Por tal razón, el control del lenguaje jurídico constituye una dura tarea, debido a la naturaleza híbrida de este lenguaje, cuyo contenido, como ya se dijo, es expresamente prescriptivo.

Xavier Linant explica que la elaboración de un tesoro para un sistema de información jurídica se alimenta por procesos tales como la investigación (servicios profesionales y el ámbito académico), las estadísticas para la coordinación política, los productos de la industria editorial especializados en la rama, y la colaboración entre la jurisprudencia (papel de los abogados) y la gestión de la información (tareas del bibliotecólogo). Nosotros agregaríamos a esto un elemento decisivo y determinante: la tecnología emergente. Sin embargo, estas relaciones son intrincadas; la única fuente que permea directamente en el tesoro son los trabajos de los legisladores, o sea, las leyes.

Por lo que la revisión de éstas es imperiosa. Un legislador puede innovar o revivir un concepto jurídico y este tendrá que tener su representatividad en el vocabulario irremediabilmente. De lo discurrido anteriormente, se puede observar en la FIGURA 1 los diversos escenarios de competencia informativa en donde interviene el legislador, quien interactúa con diferentes sistemas.

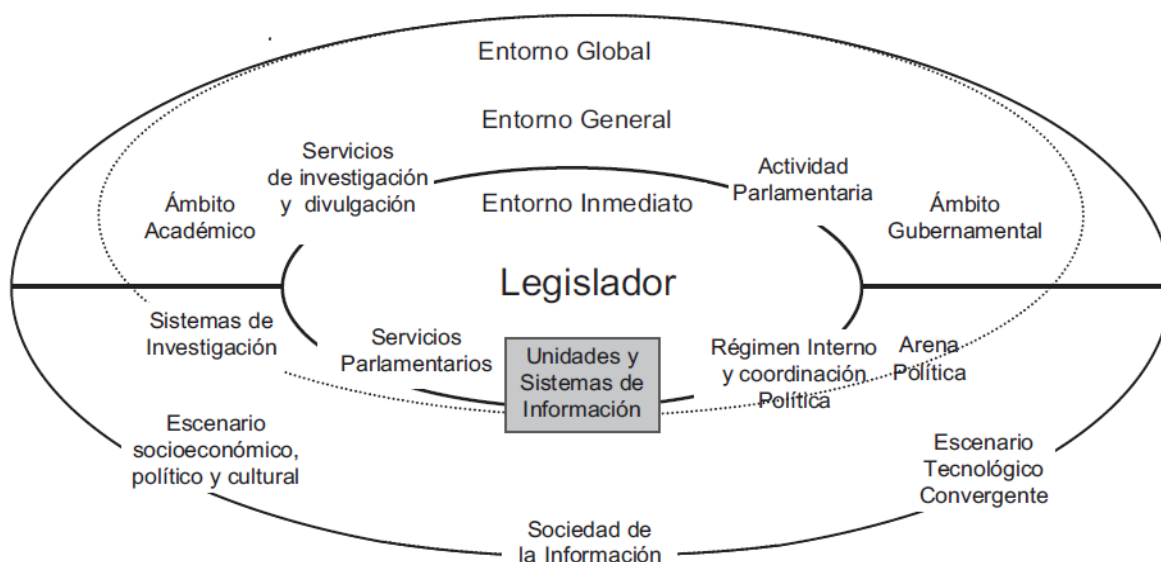


FIGURA 1

Del gráfico previo podemos aducir que la sociedad de la información, con el escenario tecnológico convergente y el socioeconómico político y cultural, engloban y determinan, a su vez, al escenario descrito anteriormente, así como a los entornos global, general e inmediato, es decir, es un macro entorno. Gran parte de este entorno se encuentra inmerso en el medio digital. Por ende, es posible la propuesta de aplicación de tesauros orientada hacia representar información usando la Web semántica, planteando la constitución de un modelo dirigido al desarrollo de aplicaciones web para la gestión de lenguajes controlados.

Las tecnologías actualmente utilizadas para la elaboración de la mayor parte de los contenidos Web carecen de estructuras que permitan una representación adecuada de los aspectos semánticos. Los motores de búsqueda utilizan técnicas basadas en la frecuencia de aparición de palabras en los procesos de recuperación, por lo que las nuevas herramientas asociadas a la Web 2.0 han hecho patentes ciertos problemas inherentes a este modelo de recuperación (Touminen, 2008). Generalmente, en los sistemas jurídicos el análisis de contenido se realiza mediante la medición del grado de

similitud semántico-sintáctica entre los diversos documentos fuente para lograr mayor confiabilidad en las respuestas en la búsqueda de información.

La Web semántica parece constituir una solución viable, ya que realiza una búsqueda similar, fundamentada en el uso de tecnologías como el lenguaje XML, los esquemas RDF y el lenguaje de marcado OWL (en su vertiente Full), en donde los metadatos y las ontologías constituyen el núcleo conceptual de desarrollo. No obstante, la siguiente declaración de Floridi debe servir como punto de inflexión para los profesionales de la información y para tomar con una postura crítica la propuesta de los tesauros informatizados para recuperar información en el entorno digital:

Demasiada retórica y muy poco de detalle hacen de la Web semántica un proyecto “conceptualmente” embrollado. Conceptos clave como “semántica”, “significado”, “comprensión”, “información”, “conocimiento” e “inteligencia”, son generosamente esparcidos en la literatura concerniente a la Web semántica, en donde parece que este calificativo es usado de manera holgada o simplemente de forma metafórica. La realidad es que los lenguajes, los protocolos y las ontologías para metadatos y metasintaxis sí pueden producir integración, sindicación y la capacidad de agregar, compartir y discernir entre datos de temas heterogéneos, pero esto por la interacción de diferentes bases de datos. Incluso en lo virtual, no parece haber nada de “semántico”. En 1997, por ejemplo, el W3C definió la primera especificación del RDF, lo cual se convirtió en una recomendación para 1999. RDF provee un lenguaje basado en tripletas de representación para identificadores Universales del Recurso (URI). Pero ninguna significación o inteligencia juega algún papel en esto (Floridi, 2009: 29).

La descripción de la semántica de los recursos de información resulta bastante compleja debido al estado de desarrollo actual de la Web, pero es posible dar un paso previo al realizar una descripción y caracterización de dichos recursos empleando metadatos y esquemas conceptuales, tales como los tesauros (Jones, 1993: 168). Un tesoro, de entrada, maneja una semántica simple tan sólo manejar una clasificación de sus descriptores, relacionándolos de forma jerárquica por generalización-especialización.

Para ello, se hace necesario definir las ventajas de los tesauros documentales y los informatizados. Ante lo cual, existen múltiples iniciativas de representación de tesauros que han dado lugar al desarrollo de SKOS, un modelo para esquemas conceptuales aplicables en la Web semántica y que actualmente se encuentra en desarrollo por el W3C.

2. El modelo de SKOS (Simple Knowledge Information System)

SKOS es una iniciativa de W3C en forma de aplicación de RDF que proporciona un modelo para representar la estructura básica y el contenido de esquemas conceptuales como listas de encabezamientos de materia, taxonomías, tablas de clasificación, tesauros o cualquier tipo de vocabulario controlado o lenguaje especializado. SKOS es un modelo que permite representar la estructura básica y el contenido de estos esquemas conceptuales, diseñado para crear nuevos sistemas de organización adaptándolos al uso en la Web semántica, ya que permite estructurar los datos en forma de tríadas que pueden ser codificadas en cualquier sintaxis válida para esquemas de RDF. En SKOS cada concepto se identifica con una cadena de caracteres de tipo URI (Uniform Resource Identifier). Estos conceptos pueden etiquetarse en cadenas de texto en uno o varios idiomas y estructurarse a través de relaciones de diversa tipología. Al respecto, Floridi nos recuerda que el idioma humano prospera al utilizar un mismo término para querer decir cosas diferentes, sin embargo esto es difícil en los procesos de automatización. Es por ello que usar a un URI diferente para cada concepto específico soluciona ese problema. El autor italiano nos indica que utilizar estos identificadores es como usar una dirección postal que puede ser distinguida de alguna que sea una dirección callejera, y ambas pueden ser distinguidas de una dirección que sea un discurso (Floridi, 2009: 28).

118

El RDF es un marco de trabajo que define a un conjunto estandarizado de conceptos y metadatos para representar la información en la web. Con RDF es posible caracterizar ciertas propiedades de conceptos de un dominio concreto al tiempo que es posible jerarquizar dichos conceptos. Aun así, es necesario alcanzar mayor grado de expresividad y abstracción para representar funciones y axiomas de una ontología. Y es que el modelo de datos SKOS es en realidad una ontología definida con OWL, que es un lenguaje de marcado para publicar y compartir datos usando ontologías en internet. Está diseñado expresamente para su uso en aplicaciones que requieran realizar funciones de procesamiento del contenido informativo y no limitarse únicamente a la presentación de la misma para su consulta por parte de personas.

SKOS puede ser utilizado conjuntamente con OWL para expresar formalmente estructuras de conocimiento sobre un dominio concreto, ya que no puede realizar esta función al no tratarse de un lenguaje para la representación de conocimiento formal (Pastor Sánchez, Martínez Méndez & Rodríguez Muñoz, 2009). La estructuración del

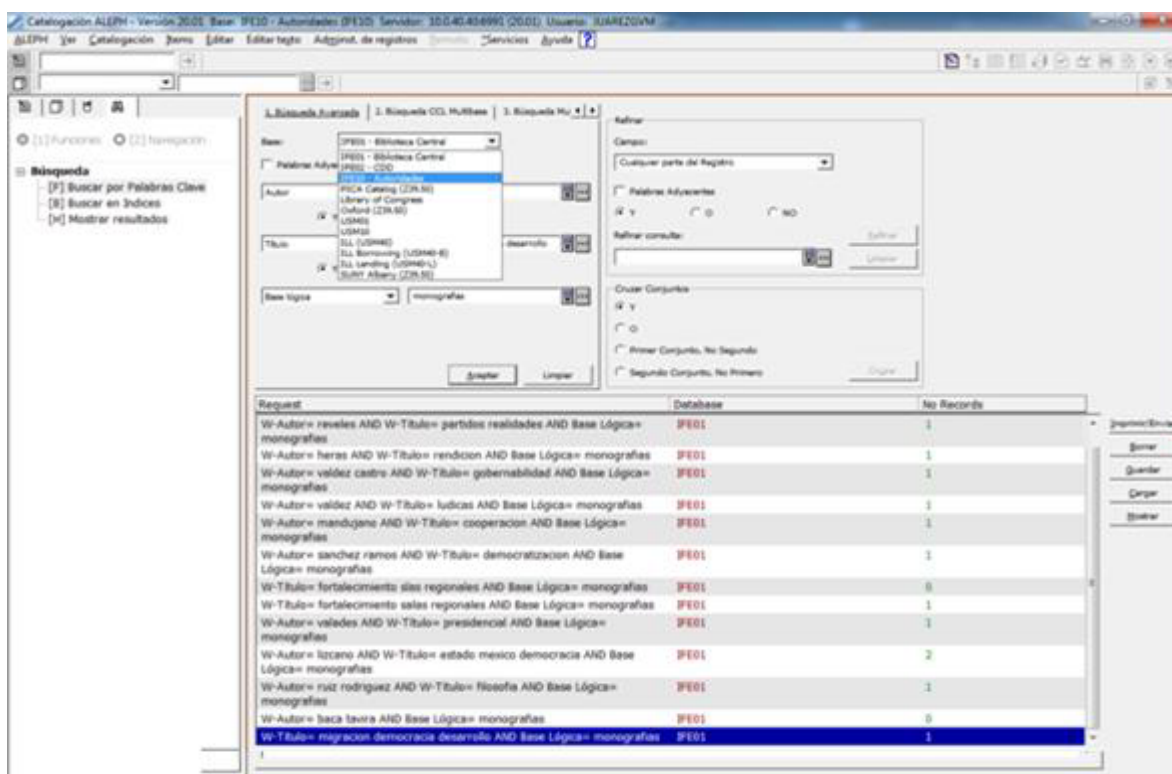
vocabulario SKOS se da en lenguaje XML, lo que permite el almacenamiento de estructuras de información en sistemas de gestión de información basados en aplicaciones Web, para el intercambio de datos entre aplicaciones y bases de datos. El vocabulario SKOS se divide en siete elementos o aspectos (TABLA 1), cada uno subdividido en etiquetas, mismas que identifican un concepto con una expresión del lenguaje natural. Un concepto se puede denotar mediante diversas etiquetas (dando expresión a los sinónimos), mientras que una etiqueta puede significar varios conceptos (dando expresión a los homónimos).

TABLA 1

Conceptos y esquemas de conceptos	skos:Concept skos:ConceptScheme skos:inScheme skos:hasTopConcept
Etiquetas léxicas	skos:prefLabel skos:altLabel skos:hiddenLabel
Relaciones semánticas	skos:semanticRelation skos:broaderTransitive skos:narrowerTransitive
La documentación	skos:note skos:ScopeNote skos:historyNote skos:changeNote skos:definition skos:editorialNote skos:example
Colecciones de conceptos	skos:Collection skos:OrderedCollection skos:member skos:memberList
Notaciones	skos:notation
Propiedades de mapeado	skos:mappingRelation skos:exactMatch skos:broadMatch skos:narrowerMatch skos:relatedMatch

3. Evaluación del software

Luego de una breve descripción del modelo SKOS es necesario evaluar varios software o paquetes que destaquen ciertas ventajas fundamentales para gestionar las relaciones y los formatos de visualización, como la capacidad de exportación de ficheros a otros sistemas. Dicha evaluación se centró en las necesidades del Instituto Federal Electoral, es decir, no se tenía contemplado un costo muy alto para la compra de un software y se necesitaba que fuera multilingüe y con una interfaz monousuario. Actualmente, el tesaurus del IFE disponible en línea está estructurado de forma peculiar bajo el módulo de Aleph (FIGURA 2). Los términos de indización fueron asentados en las Autoridades del módulo, tendente a encabezamientos de materia. Por lo tanto, más que de un tesaurus en términos propios, se cuenta con una lista de encabezamientos de materia “tesaurizados”.



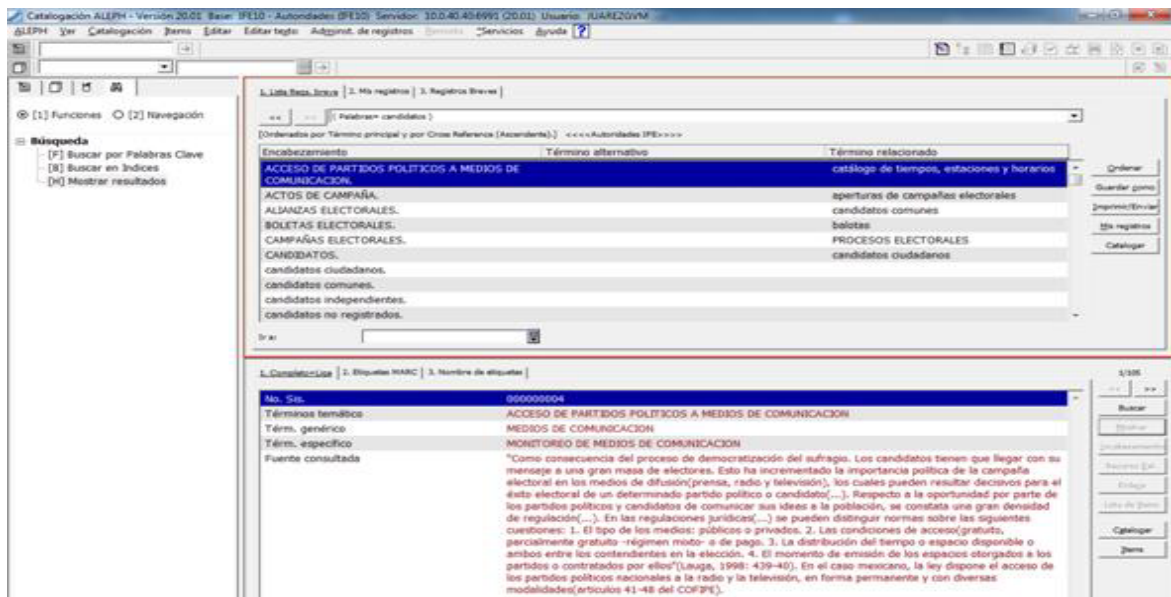


FIGURA 2

En la evaluación se consideró:

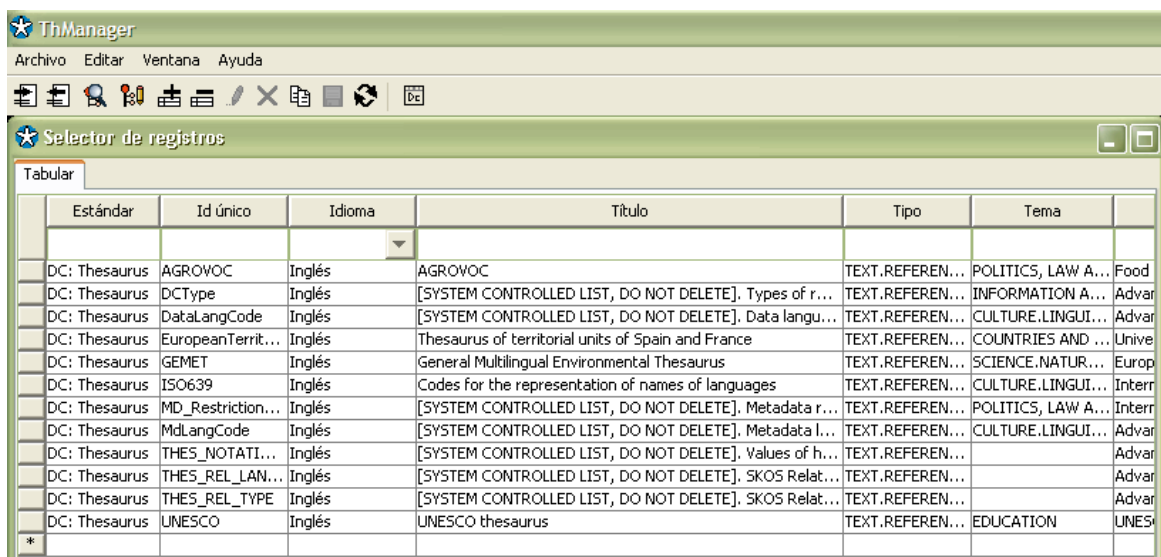
- ❖ Libre distribución
- ❖ Criterios
 - Funcionalidad
 - Herramientas conceptuales soportadas
 - Integrabilidad
- ❖ Formatos de representación de tesauros (SKOS/RDF, XML)
- ❖ Distribución como software reutilizable (paquetes, servicios)

El aspecto que tiene que ver con los formatos de representación es de suma importancia, ya que la vinculación con bases de datos e integración de información depende de esto. Tanto el lenguaje XML como el modelo SKOS fueron creados por el W3C. Básicamente, se evaluaron cuatro softwares: TemaTres, Term Tree, PoolParty y ThManager. Los primeros están basados en SQL y en las normas de la American National Standards Institute, la NISO (National Information Standards Organization), y de la ISO (International Organization for Standardization), es decir son softwares para estructurar tesauros basados en términos. PoolParty, por su parte, es un sistema de gestión para tesauros diseñado para el uso y representación semántica en la web, principalmente en las funciones de análisis de texto. Este sistema ayuda a construir y

mantener tesauros multilingües ya que proporciona una interfaz de usuario sencilla y un par de servicios semánticos (buscadores semánticos y la navegación facetada), además de que hace el uso de las especificaciones de RDF/SKOS. El principal problema con PoolParty es que es sumamente inflexible en términos de una falta de escenarios de integración cambiantes en los actuales sistemas de información jurídica. De hecho no encontramos algún tesoro de tipo jurídico estructurado con PoolParty. No obstante, sí encontramos algunos con ThManager, como el Tesoro de la UNESCO y La Nomenclatura Internacional de la UNESCO para los campos de Ciencia y Tecnología.

Pero, específicamente, en cuanto a sistemas de información jurídica, es de destacar el tesoro multidisciplinar de la Unión Europea, es decir el Tesoro multilingüe EuroVoc, el cual sirve al Parlamento Europeo y la Oficina de Publicaciones Oficiales de las Comunidades Europeas. Está estructurado en 21 campos temáticos y 127 microtesauros. Algunos descriptores de los campos temáticos 72 (los geográficos) y 76 (las Organizaciones internacionales) son multi-jerárquicos, tienen varios términos genéricos de primer nivel diferentes. Eurovoc contiene 6645 descriptores, 519 de los cuales son top terms, 7756 no descriptores, 6669 relaciones jerárquicas recíprocas y 891 notas (Alvite Díez et al., 2010: 48).

Al analizar la arquitectura de ThManager nos percatamos que diseñado para sistemas operativos Windows de 32 bits, es ligero y sencillo en cuanto al uso de sus herramientas, además que la descarga es gratuita. Es una aplicación de escritorio que requiere 100 megabytes (MB) de espacio libre en el disco duro. Con respecto a los requisitos de CPU y de memoria de acceso aleatorio (RAM), estos dependen en gran medida del tamaño y el número de tesauros cargados en la herramienta. De hecho, el software trae instalados 12 tesauros reconocidos elaborados con SKOS, entre ellos el de la UNESCO o el de AGROVOC (FIGURA 3). Este software permite reutilizar esos tesauros y modificarlos para elaborar uno nuevo o se puede comenzar uno desde un proceso en blanco. Teniendo en cuenta el número y el tamaño de tesauros utilizados como banco de pruebas, los rangos de consumo de memoria RAM pueden oscilar de 256 a 512 MB cada uno, y con un CPU de 3 Gigahertz (por ejemplo un procesador Pentium IV), los tiempos de carga para los tesauros son más que aceptables. Sin embargo, si el tamaño de los tesauros es más pequeño, se puede operar con un CPU de 1 GHz (por ejemplo un procesador Pentium III) y con 128 MB de memoria RAM.



The screenshot shows the ThManager application window. The 'Selector de registros' (Record Selector) tab is active, displaying a table of records. The table has columns: Estándar, Id único, Idioma, Título, Tipo, and Tema. The records are listed in a table with alternating light and dark rows. The last row is marked with an asterisk (*).

Estándar	Id único	Idioma	Título	Tipo	Tema
DC: Thesaurus	AGROVOC	Inglés	AGROVOC	TEXT.REFEREN...	POLITICS, LAW A...
DC: Thesaurus	DCType	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Types of r...	TEXT.REFEREN...	INFORMATION A...
DC: Thesaurus	DataLangCode	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Data langu...	TEXT.REFEREN...	CULTURE.LINGUI...
DC: Thesaurus	EuropeanTerrit...	Inglés	Thesaurus of territorial units of Spain and France	TEXT.REFEREN...	COUNTRIES AND ...
DC: Thesaurus	GEMET	Inglés	General Multilingual Environmental Thesaurus	TEXT.REFEREN...	SCIENCE.NATUR...
DC: Thesaurus	ISO639	Inglés	Codes for the representation of names of languages	TEXT.REFEREN...	CULTURE.LINGUI...
DC: Thesaurus	MD_Restriction...	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Metadata r...	TEXT.REFEREN...	POLITICS, LAW A...
DC: Thesaurus	MdLangCode	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Metadata l...	TEXT.REFEREN...	CULTURE.LINGUI...
DC: Thesaurus	THES_NOTATI...	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. Values of h...	TEXT.REFEREN...	
DC: Thesaurus	THES_REL_LAN...	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. SKOS Relat...	TEXT.REFEREN...	
DC: Thesaurus	THES_REL_TYPE	Inglés	[SYSTEM CONTROLLED LIST, DO NOT DELETE]. SKOS Relat...	TEXT.REFEREN...	
DC: Thesaurus	UNESCO	Inglés	UNESCO thesaurus	TEXT.REFEREN...	EDUCATION
*					

FIGURA 3

Inicialmente, SKOS trabaja en lenguaje JavaScript, pero a diferencia de los anteriores programas, permite transformar el tesauo a formato SKOS, por lo que se exporta el tesauo a lenguaje XML. Ciertamente, ThManager es un software de uso simple, empero, es imperioso y necesario conocer ciertas herramientas del software para diseñar un tesauo y poder exportar el contenido. Este programa se compone básicamente de tres capas: la primera, una capa de depósito donde se almacenan los tesauos y son identificados por medio de metadatos; en segundo lugar, una capa de persistencia que proporciona una interfaz de programación de aplicaciones para acceder a los tesauos almacenados en el repositorio, y, en tercer lugar, una interfaz gráfica de usuario que ofrece diferentes componentes gráficos para visualizar los tesauos, para buscarlos por sus propiedades, y para modificarlos de diferentes maneras (Lacasta, et al., 2007: 43).

A estas alturas, se hace necesario explicar cómo se estructura ese tesauo, es decir cuáles son los principios y relaciones con los que se constituye antes de ser almacenado. Es importante aclarar que, antes de introducir términos y conceptos al tesauo, el programa nos insta a elegir uno o más idiomas, es decir es multilingüe (Figura 4). Esto es importante ya que si elaboráramos el tesauo en un software en inglés aquellos términos con signos especiales (á, ñ, ü), al ser ordenados alfabéticamente, serían mandados hasta el final de la lista de términos. Varios software —aún bajo el modelo SKOS—, a pesar de la norma de los Comités Técnicos de Normalización, UNE 50125 (acrónimo de Una Norma Española), para la construcción de un tesauo multilingüe, no manejan esta opción.

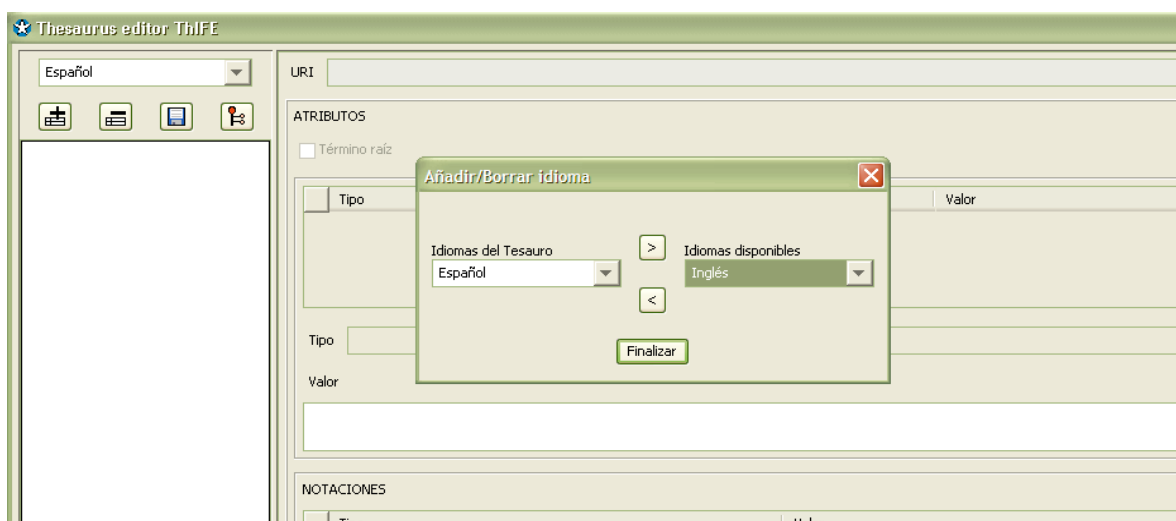


FIGURA 4

En cuanto a las interrelaciones semánticas, ThManager sólo permite relacionar términos con base en tres formas: abuelo-padre, hijo y relacionado. Esto indica, en primera instancia, que el término que se agrega está subordinado jerárquicamente a otro término, y también acepta que un término agregado tiene un término que se subordina a este. La relación del término “relacionado”, permite enlazar términos del mismo nivel, por ejemplo si un término padre tiene varios términos hijo, estos serán del mismo nivel, es decir su vínculo sería relacionado. Los términos ingresados al tesauro deberán ser descritos y relacionados en base a una serie de pasos del **Editor de tesauros** de ThManager. Los URI permiten ubicar el término en base a una línea de caracteres que se definen para el término en particular, lo que permite relacionar un término con otro a través de la asociación de los URI. Por ejemplo, para un término como Sistema jurídico su URI sería http://sistema_jurídico

124

. Bajo esta tónica, expondremos un ejemplo. El tesauro que se actualizará será el Tesauro Electoral del IFE, el cual maneja en su árbol conceptual once temas generales:

1. Conocimiento electoral: Producción teórica y su sistematización
2. Sistema político y elecciones
3. Sistemas electorales
4. Derecho y legislación electorales
5. Instituciones electorales

6. Organizaciones político-electorales
7. Sistema Jurídico
8. Geografía electoral
9. Comunicación político electoral
10. Cultura electoral
11. Historia electoral

Se elige para ejemplificar el concepto de SISTEMA JURÍDICO, que es el sétimo de la lista. Entre la ramificaciones de su primera clase se encuentra el descriptor Constituciones:


- 7 Sistema jurídico
 - 7.1 Constituciones
 - 7.1.1 Absolutismo
 - 7.1.2 Cambio constitucional
 - 7.1.3 Derecho (principios generales)
 - 7.1.4 Derecho electoral
 - 7.1.5 Derecho electoral (fuentes)

El tesauro nos insta a establecer las relaciones entre estos términos. SISTEMA JURÍDICO es el término abuelo, CONSTITUCIONES es el término padre y los puntos subsecuentes son términos hijo, y estos mantienen una relación ordenada entre sí. Esto expone el carácter de transitividad de algunas etiquetas de SKOS, principalmente las de las relaciones semánticas (Vid. TABLA 1), ya que no se puede vincular directamente SISTEMA JURÍDICO a ABSOLUTISMO, es decir de abuelo a hijo, sino que debe ser de abuelo a padre y de este a hijo.

En el caso anterior tendríamos un término abuelo, un padre y cuatro hijos (FIGURA 5) las cuales podemos revisar con el menú desplegable del **Visor del tesauro** (FIGURA 6), donde aparecen todas las relaciones establecidas en el **Editor de tesauros**: CONCEPTOS ELECTORALES tiene como término padre a SISTEMAS ELECTORALES y como término hijo a DERECHOS POLÍTICO-ELECTORALES:



FIGURA 5

Para que aparezca la interfaz anterior, es necesario precisar cómo se asientan los UP, los TG, los TE, los TR, las NA y la definición para el término CONSTITUCIONES. En primer lugar, en ThManager todos estos aspectos se asientan en el Editor, cuyo símbolo en la barra de herramientas del software es: . La definición se asienta como se observa en la FIGURA 6:

126



FIGURA 6

Para la nota de alcance, véase la FIGURA 7:

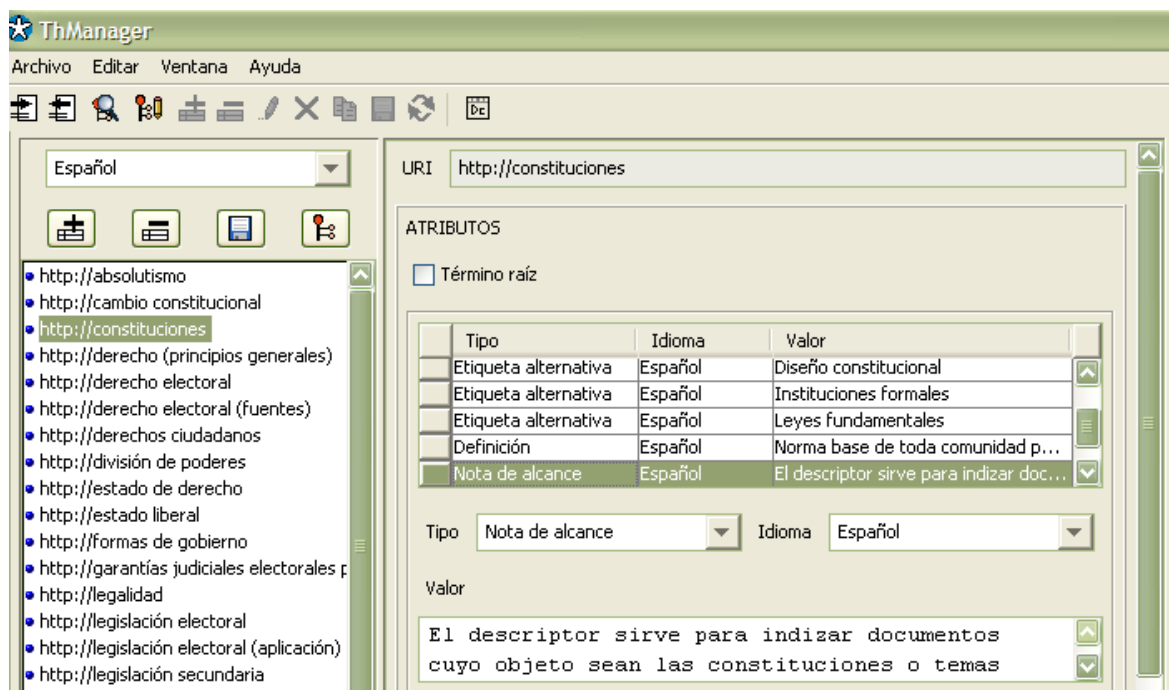


FIGURA 7

Para mirar cómo se asientan los ÚSESE POR, véase la FIGURA 8, donde se observará que se asientan como términos equivalentes y, luego, el software los toma como sinónimos (Vid. FIGURA 5):

FIGURA 8

Para los TG (abuelo o “padre del padre”), véase la FIGURA 9:

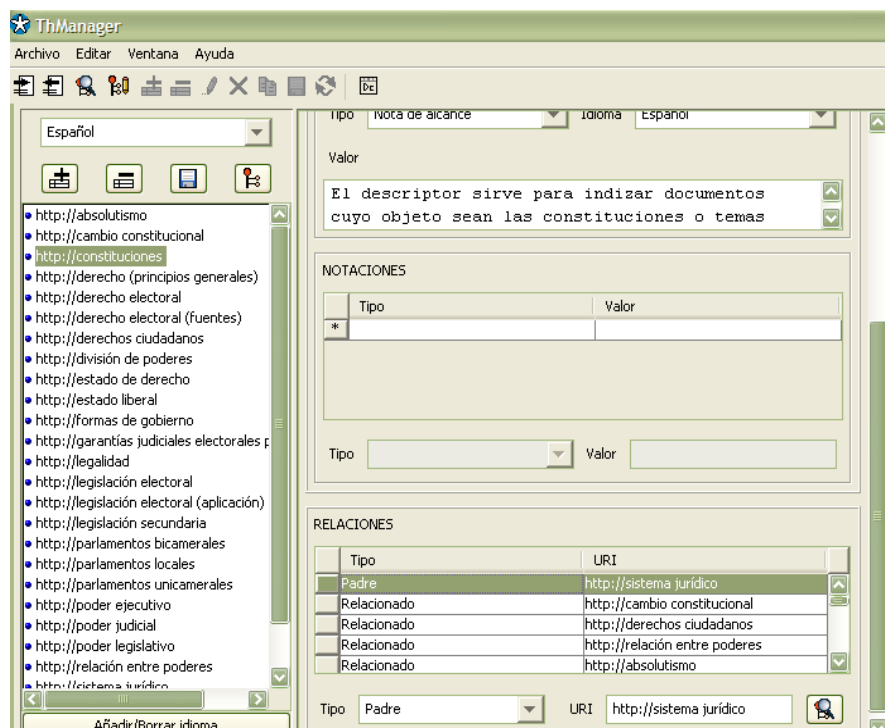


FIGURA 9

Para la gran cantidad de TE (los hijos), véase la FIGURA 10:

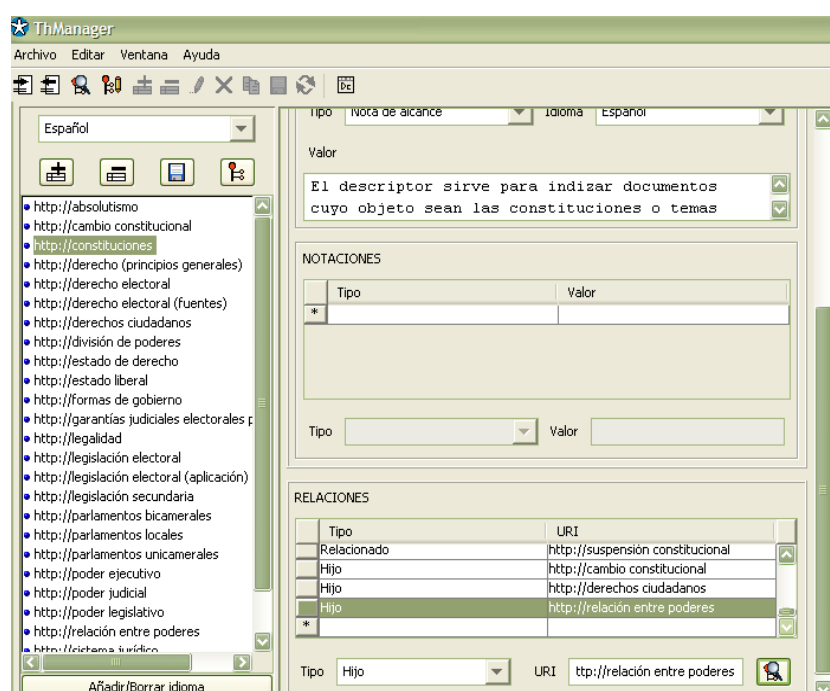


FIGURA 10

Para los TR, véase la siguiente figura:

:

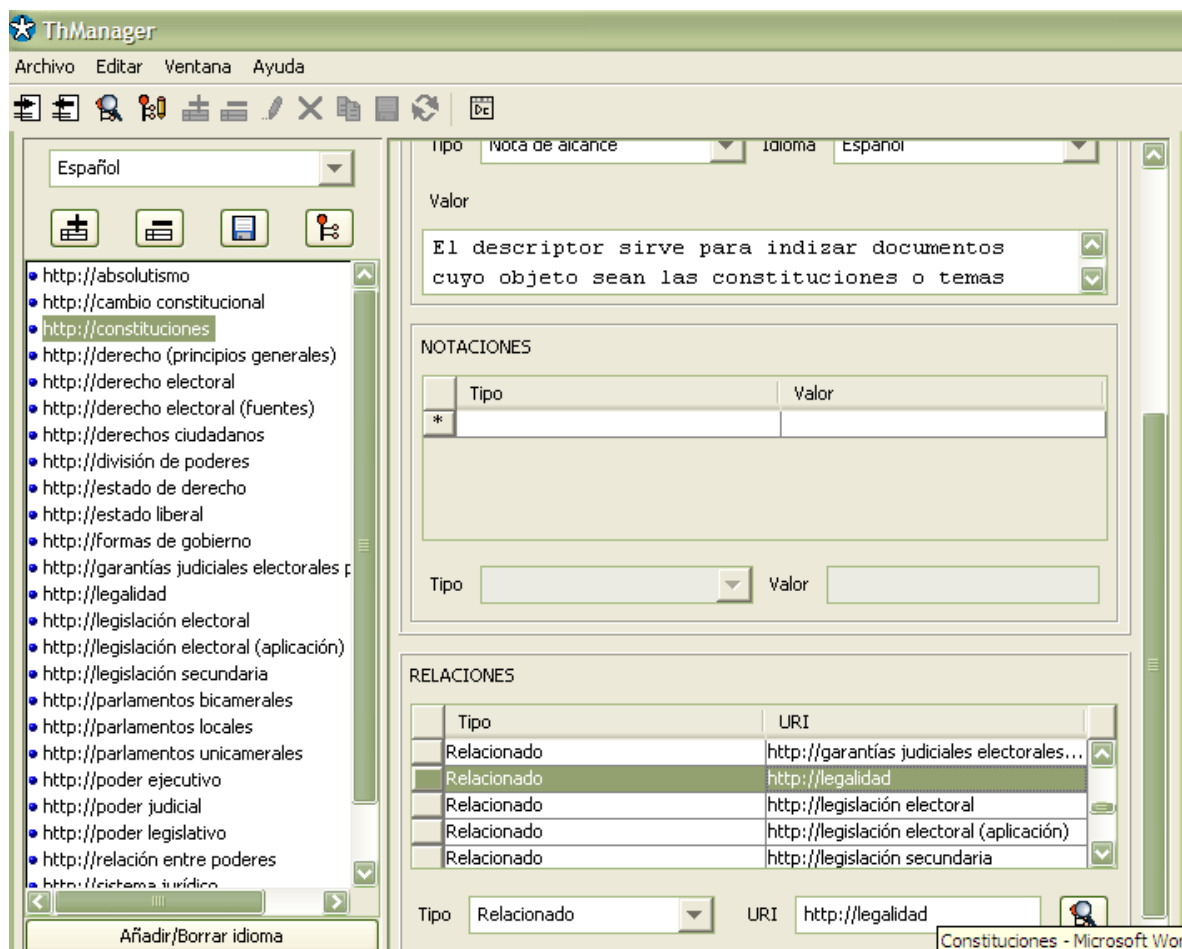


FIGURA 11

ThManager no sólo se ha creado como una herramienta independiente para facilitar la gestión de tesauros, sino también para permitir una fácil integración en las herramientas que necesitan utilizar tesauros, por lo que vincula diferentes bases de datos y repositorios con el tesauro. Por ejemplo, en una herramienta de creación de metadatos, se puede utilizar para proporcionar el componente gráfico para seleccionar los valores controlados de tesauros y automáticamente insertar en los metadatos. También se puede utilizar para proporcionar la lista de valores posibles para utilizar en un sistema de búsqueda Web, o para proporcionar una navegación basada tesauro de una colección de recursos en un sistema de búsqueda exploratoria.

Conclusiones

En este artículo se ha presentado una evaluación y descripción de las necesidades para la estructuración del Tesauro del IFE a partir de la herramienta para la gestión de los tesauros ThManager, bajo el modelo SKOS. La herramienta ThManager se ofrece como la solución a estos problemas. Es una herramienta de código abierto que puede gestionar tesauros almacenada en SKOS, lo que permite su visualización y edición. Gracias a la arquitectura de capas, sus componentes se pueden integrar fácilmente en otras aplicaciones que necesitan utilizar tesauros u otros vocabularios controlados. Adicionalmente, los componentes pueden ser utilizados para controlar los posibles valores utilizados en un servicio de búsqueda de Internet con la finalidad de facilitar búsquedas tradicionales o exploratorias basadas en un vocabulario controlado.

El modelo SKOS, y específicamente el software ThManager, son idóneos para crear tesauros de sistemas de información jurídica, en este caso, uno que atañe al derecho constitucional y a temas electorales. Debido a la naturaleza híbrida de los conceptos y mensajes legales, es necesario poder navegar en un tesauro estructurado para ser utilizado con la versatilidad que ofrece una herramienta lingüística integrada de manera computacional. Dicho en otras palabras, en un tesauro documental las relaciones semánticas se establecen entre los propios términos. A diferencia del nombrado tesauro conceptual, donde las relaciones entre conceptos suelen incluir las que desarrollan las estructuras jerárquica y asociativa, mientras por otro lado se denotan las relaciones existentes entre términos asociados a un mismo concepto (abreviaturas, uso común, tecnicismos, entre otros) e incluye una gama amplia de equivalencias entre términos, siempre vinculados con los propios documentos del sistema.

Obras consultadas

ALVITE DÍEZ, María Luisa et al. 2010. Propuesta de representación del tesauro EuroVoc en SKOS para su integración en sistemas de información jurídica. Scire: Representación y Organización del Conocimiento, 16 (2): 47-51.

CESTEURS, Werner, Smith, Barry & Goldberg, Louis. 2005. A terminological and ontological analysis of the NCI Thesaurus. Methods of Information in Medicine 44: 498-507.

DEWEZE, André. 1998. Informatique documentaire. París: Masson.

ERK, Katrin. 2009. Supporting inferences in semantic space: Representing words as regions. En Harry Bunt, Volha Petukhova y Sander Wubben (Eds.), *Proceedings of the 8th International Conference on Computational Semantics*, 1 (104-115). Tilburg, Países Bajos: Tilburg University, Department of Communication and Information Sciences.

FLORIDI, Luciano. 2009. Web 2.0 vs. the Semantic Web: A philosophical assessment. *Episteme*, 6 (1): 25-37.

JONES, Susan. 1993. A thesaurus data model for an intelligent retrieval system. *Journal of Information Science*, 19 (3): 167-178.

LACASTA, Javier et al. 2007. ThManager: An open source tool for creating and visualizing SKOS. *Information Technology and Management*, 26 (3): 39-51.

LINANT DE BELLEFONDS, Xavier. 1981. *L'informatique et le droit*. París: Presses Universitaires de France.

MATUTE CIFUENTES, Sergio Luis. 1989. Los sistemas de información, la informática jurídica y el sistema UNAM-JURE. En Jean Louis Bilon et al. (Coords), *Diálogo sobre la informática jurídica*. México: UNAM, Instituto de Investigaciones Jurídicas.

PASTOR SÁNCHEZ, Juan Antonio, Martínez Méndez, Francisco Javier & Rodríguez Muñoz, José Vicente. 2009. Advantages of thesaurus representation using the Simple Knowledge Organization System (SKOS) compared with proposed alternatives. *Information Research*, 14 (4). Documento en línea: <http://informationr.net/ir/14-4/paper422.html>.

131

PERALES OJEDA, Alicia. 1981. Traza de un sistema de información jurídica. *Anuario de Bibliotecología*, 4 (2): 7-78.

ROSALES SALINAS, Fabiola Elena. 2010. The parliamentary library as a bridge between Congress and citizenship. The Mexican Chamber of Deputies reference service. *Library Trends*, 58 (4): 510-516.

SACCO, O. & BOTHOREL, C. (2010). Exploiting semantic web techniques for representing and utilising folksonomies. En *Proceedings of the international workshop on modeling social media*. Nueva York: ACM.

TÉLLEZ VALDÉS, Julio Alejandro. 2008. *Derecho informático*. México: McGraw-Hill

TUOMINEN, Jouni et al. 2008. ONKI-SKOS - Publishing and utilizing thesauri in the semantic web. En *Proceedings of the 13th Finnish Artificial Intelligence Conference (STeP 2008)*: Finnish Artificial Intelligence Society. Espoo, Finlandia: J. V. Editors.

TUDHOPE, Douglas, HARITH, Alanis & JONES, Christopher. 2001. Augmenting thesaurus relationships: Possibilities for retrieval. *Journal of Digital Information*, 1 (8): 1-20. Documento en línea: http://eprints.soton.ac.uk/254484/1/Tudhope_JoDI.pdf.